DOCUMENT RESUME

ED 197 645 HE 013 411

AUTHOR Marsh, Herbert W.
TITLE Students' Evaluations of College/University Teaching: A Description of Research and an Instrument.
PUB DATE 80
NOTE 35p.: Not available in paper copy due to marginal legibility of original document. Earlier version of this paper was presented at the Annual Meeting of the Australian Association for Research in Education (Sydney, Australia, November 1980).

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Evaluation Methods: Factor Analysis: *Faculty Evaluation: Higher Education: *Measurement Techniques: *Questionnaires: Self Evaluation (Individuals): *Student Evaluation of Teacher Performance: *Teacher Effectiveness: Test Validity
IDENTIFIERS *Students Evaluation of Educational Quality

ABSTRACT

Research that led to the development of an instrument and a program for collecting students' evaluations of college/university teaching is summarized. The current version of Students' Evaluation of Educational Quality (SEEQ), which is appended, was developed at the University of Southern California (USC). (The first SEEQ was developed at the University of California, Los Angeles.) The one-page evaluation instrument contains standardized questions, allows for an individual instructor or academic unit to design items, and provides for student comments to open-ended questions. Research indicates that SEEQ measures nine distinct components of teaching effectiveness that have been identified in both student ratings and faculty self-evaluations of their own teaching. Reliability is good when based upon 10 to 15 or more student responses. The ratings have successfully been validated against the retrospective ratings of former students, student learning as measured by objective examination, affective course consequences, and faculty self-evaluations of their own teaching effectiveness. Feedback from student ratings, particularly when coupled with a candid discussion with an external consultant, produced improvement in both subsequent ratings and student learning. Data derived from factor analysis, a sample summary report returned to individual faculty and department heads, and references are included. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED197645

Students' Evaluations of College/University Teaching: A Description of Research and an Instrument

Herbert W. Marsh[1]

University of Sydney,

Department of Education

Australia

BEST COPY AVAILABLE

U S DEPARTMENT OF HEALTH EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Herbert W. Marsh

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Running Head: Students' Evaluations

1 Grateful acknowledgements are extended to John Schutz, Joseph Kertes, and Robert Linnell of the University of Southern California, and Raymond Orbach, James Trent, Robert Pace, and Leon Levine of the University of California, Los Angeles, who supported the student evaluation program at their universities. Thanks are also extended to each of the coauthors of studies that led to the development of SEEQ, and particulary to Jesse Overall. The author was employed as Director of the Evaluation of Instruction Program at UCLA (1972-74) and as Head of Evaluation Services at USC (1976-80) during the time this research was being conducted. An earlier version of this paper was presented at the annual meeting of the Australian Association for Research in Education, Sydney, Australia, November, 1980. Requests for reprints should be sent to Herbert W. Marsh, Department of Education, University of Sydney, Sydney, NSW 2006, Australia.

HE 013 411

Abstract

The purpose of this study is to describe research that led to the development and implementation of SEEQ (Students' Evaluations of Educational Quality). SEEQ is an instrument and program for collecting students' evaluations of college/university teaching. Research to be described indicates that SEEQ measures nine distinct components of teaching effectiveness that have been identified in both student ratings and faculty self-evaluations of their own teaching. Reliability is good when based upon 10 to 15 or more student responses. The ratings have successfully been validated against the retrospective ratings of former students, student learning as measured by objective examination, affective course consequences, and faculty self evaluations of their own teaching effectiveness. Suspected sources of bias to the ratings have been shown to have little impact. Feedback from student ratings, particularly when coupled with a candid discussion with an external consultant, produced improvement in both subsequent ratings and student learning.

Students' Evaluations of College/University Teaching : A Description of Research and an Instrument

Background

The purpose of this review is to summarize research that led to the development of SEEQ (Students' Evaluations of Educational Quality). SEEQ is an instrument and a programme for collecting students' evaluations of college/university teaching. Research presented in this review is described in greater detail in a series of technical reports and publications. This research, in addition to guiding SEEQ's development, has also provided an academic credibility that is essential in winning faculty support. It is hoped that this review may serve as both a model and encouragement to academic units seeking to implement or improve systematic programs of students' evaluations.

Research and development on the first SEEQ, which is substantially similar to the current version, was conducted at the University of California, Los Angeles (UCLA). This effort began with a Task Force on the Evaluation of Teaching that examined evaluation practices at UCLA and other universities, and made recommendations that included the development of a campus-wide program of students' evaluations of teaching. Based upon current practices, interviews with students and faculty, and a review of the evaluation literature, an extensive item pool was developed. The work done by Hildebrand, Wilson and Dienst (1971) at the University of California, Davis was particularly important in developing this pool of items. Several different pilot surveys - each consisting of 50-75 items - were administered to classes in different academic departments. Students, in addition to making ratings, were asked to indicate the items they felt were most important in describing the quality of teaching. Similarly, faculty were asked to indicate the items they felt would provide them with the most useful feedback about their

teaching. Students' open-ended comments were reviewed to determine if important aspects had been excluded. Factor analysis identified the dimensions underlying the student ratings, and the items that best measured each. Reliability coefficients were compiled for each of the evaluation items. Finally, after several revisions, four criteria were used to select items to be included on the UCLA version of SEEQ. These were: 1) student ratings of item importance, 2) faculty ratings of item usefulness, 3) factor analysis, and 4) item reliabilities. During the last six years over 500,000 of these forms have been completed by UCLA students from more than 50 academic departments in over 20,000 courses. The results of the evaluations are returned to faculty as feedback about their teaching, are used in tenure/promotion decisions, and are published for students to use in the selection of courses.

The current version of SEEQ (see Appendix I) was developed at the University of Southern California (USC). A preliminary version of the instrument was adopted on a trial basis by the Division of Social Sciences, pending the outcome of research on the instrument. On the basis of much of the research summarized in this review, the current form was unanimously endorsed by the Dean and Department Chairpersons in the Division, and its use required in all Social Science courses. The program was later adopted by other academic units at USC, and over 250,000 SEEQ forms have been completed by USC students over the last four years.

Description of the Instrument

The SEEQ survey form is presented in Appendix I. The one-page evaluation instrument is self-explanatory, easily administered, and computer scorable. The form strives for a compromise between uniformity and flexibility. The standardized questions used to evaluate all courses measure separate components of instructional effectiveness that have been identified with factor

analysis. Provision for Supplemental Questions allow the individual instructor or academic unit to design items unique to specific needs. Adequate provision for student comments to open-ended questions is provided on the back of the form.

A sample of the two-page summary report prepared for each course is presented in Appendix II (the actual report appears on 8.5 inch x 15 inch computer paper). The summary report, along with the completed surveys that contain students' open-ended comments, are returned to the instructor. Copies of the report are also sent to the Department Chairperson and/or the Dean of the particular academic unit. The data upon which the report is based is permanently stored in a computer archive system by the Office of Institutional Studies, the central office that processes the forms. In the report, the evaluation factor scores, the overall summary ratings, and demographic/ background items are presented on page 1, while the separate rating items appear on page 2. Each item is summarized by a frequency distribution of student responses, the mean, the standard error, and the percentile rank that shows how the mean rating compares with other courses. A graphic representation of the percentile rank is also shown. If any supplemental questions were used, a summary of these responses appears on a third page.

The normative comparisons provided in the summary report (the percentile ranks) play an important role in the interpretation of the ratings. First, students are universally quite generous in their evaluations of teaching. The average overall course and instructor ratings are typically about 4.0 on a one-to-five scale. Second, some items receive higher responses than do others - overall instructor ratings are almost always higher than overall course ratings. Finally, comparisons are made between instructors teaching courses at similar levels (i.e., there are separate norms for graduate level courses, undergraduate level courses taught by faculty, and courses taught by teaching assistants). Academic units at USC (e.g., the 10 departments in the Division of Social Sciences) are given the option of using university-wide norms or norms based upon ratings from just their unit. However, ratings are only ranked against norms containing at least 200 courses.

A longitudinal summary report, summarizing all the available courses ever taught by each instructor is also produced annually. The report contains means and percentile ranks for the evaluation factor scores, the overall summary ratings, and selected background/demographic items. This information is presented separately for each course, and is averaged across all graduate level courses and across all undergraduate courses. Courses that were evaluated by fewer than 10 students or by less than 50% of the enrolled students are not included in the longitudinal averages. Important information can be gained from examining this report, beyond the convenience of having a summary of all the ratings for each teacher. The longitudinal average is not unduly affected by a chance occurrence in any one course offering, and it reflects teaching effectiveness in the range of courses that are likely to be taught by a particular instructor. The change in ratings over time provides a measure of instructional improvement. Furthermore, this summary provides a basis for determining the classes in which an individual teacher is most effective.

In addition to the individual and longitudinal summary reports, other studies and special analyses are performed at the request of the Dean and/or Chairpersons. These include requests as diverse as using previous ratings for a particular course as a baseline against which to compare ratings after an innovative change, a determination of the trend over time in ratings of all courses within a given academic department, and the use of supplemental questions to query students about their preferences in class scheduling.

Summary of Research

Factor Analysis

Factor analysis is used to describe the different components of teaching effectiveness actually being measured by a set of questions. Its use

is particularly important in the development of student evaluation instruments, since it provides a safeguard against a "halo effect" - a generalization from some subjective feeling about the teacher which affects ratings of all the questions. To the extent that all the items are contaminated by this halo effect, they will all blend together and not be distinguished as the separate components of teaching effectiveness that the evaluation form was designed to measure.

A well developed factor structure is also important to the interpretation of the student ratings. Broad global ratings averaged across a collection of heterogeneous items provide little diagnostic feedback and are difficult to interpret. For example, Marsh, Overall and Kesler (1979b) showed that while large classes did tend to receive lower ratings when averaged across all items, this effect was limited almost entirely to the Group Interaction and Individual Rapport factors. Similarly, an interview with a student about an earlier version of the evaluation form indicated that she had given an instructor lower ratings on several more or less randomly selected items because there were no items where she could express her sentiment that "the examinations were terrible". Even if particular components of teaching effectiveness seem less important to a particular instructor (or academic unit), their exclusion will make other ratings more difficult to interpret.

SEEQ measures nine evaluation factors (see Table 1). Marsh (Marsh and Overall, 1979b; Marsh, In Press) presented a factor analysis of student ratings that confirmed the nine factors SEEQ was designed to measure, and these findings have been replicated in different academic disciplines and in different academic years. Even more convincing support came from a study in which faculty in 329 classes were asked to evaluate

their own teaching effectiveness with the same SEEQ form that was used by their students. Separate factor analyses of the student ratings and the instructor self-evaluation both demonstrated the same nine evaluation factors that had previously been identified (see Table 1). These analyses illustrate the replicability of the rating factors and their generalizability across different methods of evaluation.

Insert Table 1
about here

Factor scores derived from the results of factor analytic research are an important part of the summaries of the student ratings described earlier. Research described in this section is presented in more detail in Marsh, Overall & Kesler (1979), Marsh & Overall (1979b), Marsh & Cooper (in press) and Marsh (in press). Further discussion of this issue is presented in Marsh (1980b).

Reliability

Reliability refers to the relative lack of random error in student ratings, and is a necessary prerequisite for any measurement device. Reliability is assessed by determining the consistency or stability of a measure. According to one conceptualization of reliability called the interclass correlation, a reliable item is one in which there is agreement among ratings within each class, but consistent differences between the ratings of different classes. A similar approach would be to take a random half of the students' ratings from each of a large number of classes and to correlate their ratings with those of the remaining students. The reliability of a given item depends more on the number

of students responding than on the actual item content. The average reliability of SEEQ items is about .90 when based upon 25 students, but falls to .74 when based upon only 10 responses and is even lower for fewer responses.

An alternative determination of reliability, called coefficient alpha, considers the relative agreement among different items designed to measure the same factor. This approach does not include disagreement among students within the same class as a source of unreliability, and probably results in an inflated estimate of reliability. The coefficient alphas for the different evaluation factors in SEEQ vary between .88 and .97.

As a consequence of this research, a caution appears on any summary report that is based upon fewer than 10 responses. Similarly, these courses are not included in the compuatation of the longitudinal averages. Data on the reliability of SEEQ items and factors is presented in Marsh & Overall (1979b).

Long Term Stability

A common criticism directed at student ratings is that students do not have an adequate perspective to recognize the value of instruction at the end of a class. According to this argument, students will only recognize the value of teaching after being called upon to apply the course materials in further coursework and after graduation. A rather unique opportunity to test this notion arose at a California State University which had adopted an earlier version of SEEQ. Undergraduate and graduate students in the school of management evaluated teaching effectiveness at the end of each course. However, unlike most programs, the forms were actually signed by the students, allowing the identification of individual responses. One year after graduation from the program (and several years after taking a course) the same students were again asked to make

"retrospective ratings" of teaching effectiveness in each course, using a subset of the original items. Since all evaluations were signed, the end-of-term ratings could be matched with the retrospective ratings. Over a several year period of time, matched sets of ratings-- both end-of-term and retrospective-- were collected for students in 100 classes. Analysis of the two sets of ratings showed remarkable agreement. The average correlation (relative agreement) between end-of-term and retrospective ratings was .83. Mean differences between the two sets of ratings (absolute agreement) was small; the median rating was 6.63 for retrospective ratings and 6.61 for end-of-term ratings. Separate analysis showed these results to be consistent at both the graduate and undergraduate levels, and across different course types.

This research is described in more detail in Marsh & OVerall (1979a) and Overall & Marsh (1980b). In related research,Marsh showed that responses from graduating seniors were similar to the ratings of current students.

Validity -- Student Learning

Student ratings, one measure of teaching effectiveness, are difficult to validate since there is no universal criterion of effective teaching. Consequently, using an approach called construct validation, student ratings have been related to other measures that are assumed to be indicative of effective teaching. If two measures that are supposed to measure the same thing show agreement, there is evidence that both are valid. Clearly this approach requires that many alternative validity criteria be used. Within this framework, evidence of the long-term stability of student ratings can be interpreted as a validity measure. However, the most

commonly used criterion has been student learning as measured by performance on a standardized examination.

Methodological problems require a special setting for this research. Ideally, there are many sections (i.e. different lecture groups that are part of the same course) of a large multi-section course in which students are randomly assigned or at least enroll without knowledge of who will be teaching the section. Each section of the course should be taught by a separate teacher, but the course outline, textbooks, course objectives, and most importantly the final examination, should be developed by a course director who does not actually lecture to the students. In two separate studies applying this methodology, it was found that the sections that evaluate teaching most favourably during the last week of classes also perform best on the standardized examination given to all sections the following week. Since students did not know who would be teaching different sections at the time of registration, and sections did not differ on a pretest administered at the start of the term, these findings provide good support for the validity of student ratings.

In the second of these studies a set of affective variables was also considered as a validity criterion. Since the course was an introduction to computer programming, these included such variables as feelings of course mastery, plans to apply the skills that were gained from the course, plans to pursue the subject further, and determination of whether or not students had joined the local computer club. In each case, more favourable responses to these items were correlated with more favourable evaluations of the teacher.

These two studies are described in more detail in Marsh, Fleiner & Thomas (1975) and Marsh & Overall (1980). Similar findings, using this same methodology, are presented in Frey, Leonard, & Beatty (1975), and Centra (1977), and in studies reviewed by McKeachie (1979) and Marsh (1980b).

Validity -- Faculty Self Evaluations

Validity research such as that described above, while supporting the use of student ratings, has generally been limited to a specialised setting (e.g. large multisection courses) or has employed criteria (e.g. student retrospective ratings) that are unlikely to convince sceptics. Thus, faculty will continue to question the usefulness of student ratings until validity criteria that are both convincing and applicable across a wide range of courses are utilized. Faculty self-evaluations of their own teaching is one criterion that meets both these requirements. Furthermore, instructors can be asked to evaluate their own teaching along the same dimensions employed in the student rating form, thereby testing the specific validity of the different rating factors. In two different studies faculty were asked to evaluate their own teaching with the same evaluation form used by their students, as well as to provide background/demographic information and to express their attitudes toward the evaluation of teaching. A letter from the Dean of the Division was also sent that encouraged participation and guaranteed confidentiality.

A majority of the faculty (59%) indicated that some measure of teaching effectiveness should be given more emphasis in promotional decisions. Faculty clearly agreed that student ratings were useful to the faculty themselves as feedback, and a majority of them even agreed that the ratings should be made publicly available for students to use

in course selection. However, they were more sceptical about the accuracy of the student ratings and even more critical of using classroom visitation or faculty self evaluations in promotional decisions; they were somewhat less critical about colleague examination of course outlines, reading lists, and classroom examinations. Faculty also indicated a number of potential biases that they felt would substantially affect student ratings. The most frequently mentioned were Course Difficulty, Grading Leniency, Instructor Popularity, and Student Interest in the Subject Before Taking the Course. A dilemma clearly exists. Faculty are concerned about teaching effectiveness, even to the extent of wanting it to play a more important role in their own promotions. However, many expressed doubts about each of the proposed measures of teaching effectiveness -- including student ratings. Before the potential usefulness of the student ratings can be realized, faculty and administrators have to be convinced that student ratings are valid.

In the first study, only undergraduate courses taught by faculty were considered. Despite faculty reservations about the validity of the student ratings, there was considerable student-faculty agreement in the evaluations of teaching effectiveness. Validity coefficients, the correlation between student and faculty ratings on the same factors, were statistically significant for all evaluation factors (median $r = .49$). Absolute agreement was also assessed by examining the mean differences between student and faculty self evaluations. Across all the evaluation items the median rating was the same for both students and faculty - 4.07 - and few differences in either direction reached statistical significance.

In the second study, the same general findings were replicated with a larger sample (329 classes) that included graduate level courses and courses taught by teaching assistants (see Table 2). Student evaluations correlated with instructor self evaluations in courses taught by teaching assistants ($r = .46$), in undergraduate courses taught by faculty ($r = .41$), and even in graduate level courses ($r = .39$), demonstrating their validity at all levels of teaching. Furthermore, a multitrait-multimethod analysis (Campbell & Fiske, 1959) also provided evidence for the distinctiveness of each of the rating factors. For example, if a single "generalized rating factor" underlies both student and instructor ratings, then agreement on any particular factor might be a function of this generalized agreement and not have anything to do with the specific content of the factor being considered. However, if this were the case, the correlations between student and instructor ratings on <u>different</u> factors should be nearly as high as correlations between ratings on the <u>same</u> factors.

In fact, while correlations between student and instructor ratings on the same factors were high (median $r = .45$), correlations between their ratings on different factors were low (median $r = .02$). This argues for the distinctiveness of the different evaluation factors and for the use of multifactor evaluation instruments that have been developed with the use of factor analytic techniques. The findings of these two studies provide further evidence for the validity of the student ratings, suggest the possible usefulness of faculty self evaluations, and should be particularly helpful in reassuring faculty about the accuracy of the student ratings.

Insert Table 2 about here

The results of the original study appear in Marsh, Overall & Kesler (1979a), while the findings of the second study are presented in Marsh & Overall (1979b), Marsh & Cooper (in press), and Marsh (in press).

Relationship with Student, Course & Instructor Characteristics

It is often feared that variables unrelated to teaching excellence may affect student ratings, and the harshest critics even suggest that faculty can "buy" favourable ratings by teaching only small courses, giving high grades, and requiring little work by students. While these attitudes are probably not held by a majority of the faculty, results cited earlier suggest that many do feel that student ratings are biased. The study of possible biases is complicated by a number of problems. First is the question of how large a relationship must be before it is considered practically significant. Second is the problem of how to interpret a relationship even if it is substantial. There are generally several alternatives and a bias may not be the most likely. For example, the positive relationship between student ratings and student learning supports the validity of the ratings, and it is unreasonable to say that student ratings are biased by student learning. While the question is complex, the first step is to determine which variables are substantially related to student ratings.

The relationship between student evaluations of 511 courses and a set of 16 student/course/instructor characteristics was examined. The set of background variables included such things as Class Size, GPA, Teacher Rank, Reason for Taking the Course, Class Level, Year in School, Expected Grade, Workload/Difficulty, and Prior Student Interest in the Subject. Separately, each background variable generally explained less than 5% of the variance

in any of the student evaluation factors, and there was little indication of nonlinearity (see Table 3). The only variable that consistently demonstrated nonlinearity was Class Size -- the smallest and largest classes tended to be rated most favourably. Several multivariate techniques showed that 12-14% of the variance in the student ratings could be explained by the entire set of background variables. Three background variables were most influential; more favourable ratings were correlated with higher Prior Subject Interest, Higher Expected Grades, and higher levels of Workload/Difficulty. A path analysis showed that Prior Subject Interest was most important, and also accounted for one-third of the relationship between Expected Grades and ratings.

Insert Table 3 about here

These results show that even the combined effect of the entire set of background variables has only a small impact on student ratings, but indicated that three of these background variables were most influential -- Workload/Difficulty, Prior Subject Interest, and Expected Grades. Although Workload/Difficulty is often suggested as a potential bias, the relationship found in this study was the opposite of the suggested bias. Harder, more difficult courses that require more time outside of class receive more favourable ratings.

Prior Subject Interest, the variable with the largest impact on ratings, was examined in greater detail in a separate study. A similar pattern of correlations was found between Prior Subject Interest and student ratings collected at both UCLA (using the earlier version of SEEQ) and USC. Prior Subject Interest was most highly correlated with ratings of Learning/Value in both settings. The relationship between Prior Subject Interest and Instructor self-evaluations was also explored in that study. Prior Subject Interest --

measured by both student and instructor perceptions -- illustrated a similar pattern of correlations with both student ratings and instructor self-evaluations. In particular, Prior Subject Interest was most highly correlated to both student and instructor ratings of Learning/Value. These findings argue that faculty actually are more effective at teaching when working with motivated students, and that this more effective teaching is accurately reflected in the student ratings.

The relationship between student ratings and Expected Grades is subject to several alternative interpretations. First, the Expected Grade effect was reduced by one-third by controlling for Prior Subject Interest. The best explanation is that Prior Subject Interest caused both better grades and a better educational experience. According to this interpretation, part of the Expected Grade relationship with student ratings is spurious. Second, the Expected Grade relationship can only be considered a bias if higher grades reflect "easy grading" on the part of the teacher. If the higher grades reflect better student achievement, then the Expected Grade relationship may support the validity of the student ratings -- better ratings are associated with more student learning. At least two facts support this interpretation. First, Prior Subject Interest is related to Expected Grades and it is more reasonable to assume that it affects student achievement rather than the instructor's grading standards. Second, faculty self evaluations of their own grading standards showed little correlation with student ratings. In reality, Expected Grades probably reflect some unknown combination of both "easy grading" and student achievement. However, even if Expected Grades do represent a real bias to the student ratings, their effect is not substantial.

These studies show that none of the suspected biases to student ratings seem to actually have much impact. Similar findings have been reported by Hildebrand, Wilson & Dienst (1971), McKeachie (1979), Marsh (1980b), and Remmers (1963). Nevertheless, as a consequence of this research,

summary reports describing student evaluations also include mean responses and percentile ranks for Prior Subject Interest and Expected Grades (see Appendix II). This research is described in greater detail in Marsh (1978, 1980b). Separate studies have examined the relationship between student ratings and: 1) Expected Grades (Marsh, Overall & Thomas, 1976), 2) Class Size (Marsh, Overall & Kesler, 1979b), and 3) Prior Subject Interest (Marsh & Cooper, in press). In related research, Marsh & Overall (in press) demonstrated that student ratings are primarily a function of the instructor doing the teaching, and not the particular course or the level at which it is taught.

Instructional Improvement -- Feedback from Student Ratings

There is ample reason to believe that a carefully planned program of instructional evaluation instituted on a broad basis will lead to the improvement of teaching. Teachers, particularly those who are most critical of the student ratings, will have to give more serious consideration to their own teaching in order to consider the merits of an evaluation program. The institution of the program and the clear endorsement by the administrative hierarchy will give notice that quality of teaching is being taken more seriously -- an observation that both students and faculty will be likely to make. The results of the student ratings -- as one measure of teaching effectiveness -- will provide a basis for administrative decisions and thereby increase the liklihood that quality teaching will be recognized and rewarded. The social reinforcement of getting favourable ratings will provide added incentive for the improvement of teaching, even at the tenured faculty level. Finally, the diagnostic feedback from the student ratings may provide a basis for instructional improvement. As described earlier, faculty at USC indicate that student ratings are useful in the improvement of a course and/or the quality of their teaching; 80% said that they were potentially useful while 59% said

they actually had been useful. However, the empirical demonstration of this suggestion is more difficult to test.

In two different studies the effect of feedback from midterm evaluations on end-of-course criteria was tested. Both these studies were conducted with the multi-section course in computer programming described earlier. In the first study, students completed an abbreviated version of the student evaluation instrument at midterm, and the results were returned to a random half of the instructors. At the end of the term, student ratings of "perceived change in instruction between the beginning of the term and the end of the term" were significantly higher for the feedback group, as were ratings on two of the seven evaluation factors. Ratings on the overall course and instructor summary items did not differ, nor did student performance on the standardized final examination given to all students.

Several changes were made in the second study that was based upon 30 classes. First, midterm evaluations were made on the same evaluation form that was used at the end of the course. Second, the researchers actually met with the group of randomly selected feedback instructors to discuss the ratings. At this meeting the teachers discussed the evaluations with each other and with the researchers, but were assured that their comments would remain confidential. A third change was the addition of affective variables, items that focused on application of the subject matter and student plans to pursue the subject. At the end of the term, students of the Feedback instructors: 1) rated teaching effectiveness more favourably, 2) averaged higher scores on the standardized final examination, and 3) experienced more positive affective outcomes than students whose instructors received no feedback. Students in the

Feedback group were similar to the other students in terms of both pretest achievement scores completed at the start of the term and the midterm evaluations of their teachers. These findings suggest that the feedback from student ratings, coupled with a frank discussion of their implications with an external consultant, can be an effective intervention for improving teaching effectiveness.

The details of these studies have been described in two published articles (Marsh, Fleiner & Thomas, 1975; Overall & Marsh, 1980).

Summary

In summary, research described in this study has indicated that:

1. SEEQ measures nine distinct components of teaching effectiveness as demonstrated by factor analysis. Factor analysis of faculty evaluations of their own teaching resulted in the same factors. Factor scores based upon this research are used to summarize the student ratings that are returned to faculty.

2. Student evaluations are quite reliable when based upon the responses of 10 to 15 or more students. Class ratings based upon fewer than 10 student responses should be interpreted carefully.

3. The retrospective ratings of former students agree remarkably well with the evaluations that they made at the end of a course.

4. Student evaluations show moderate correlations with student learning as measured by a standardized examination and with affective course consequences such as application of the subject matter and plans to pursue the subject further.

5. Faculty self evaluations of their own teaching show good agreement with student ratings.

6. Suspected sources of bias to student ratings have little impact.

7. Feedback from student ratings, particularly when coupled with a candid discussion with an external consultant, can lead to improved teaching.

References

Centra, J.A. Student ratings of instruction and their relationship to student learning. American Educational Research Journal, 1977 14, 17-24.

Frey, P.W., Leonard, D.W. & Beatty,W.W. Student ratings of instruction: Validation research. American Educational Research Journal, 1975, 12 327-336.

Hildebrand, M., Wilson, R.C., & Dienst, E.R. Evaluating University Teaching. Berkeley: Center for Research and Development in Higher Education, University of California, Berkeley, 1971.

Marsh, H.W. The validity of students' evaluations: Classroom evaluations of instructors independently nominated as best and worst teachers by graduating seniors. American Educational Research Journal, 1977, 14, 441-447.

Marsh, H.W. Students' Evaluations of Instructional Effectiveness: Relationship to Student, Course, and Instructor Characteristics. Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, March, 1978. (ERIC Document Reproduction Service No. ED 155 217.

Marsh, H.W. The influence of student, course and instructor characteristics on evaluations of university teaching. American Educational Research Journal, 1980, 17, 219-237. (A)

Marsh, H.W. Research on students' evaluations on teaching effectiveness. Instructional Evaluation, 1980, 4, 5-13 (B)

Marsh, H.W. Validity of students' evaluations of college teaching: A multi-trait-multimethod analysis. Journal of Educational Psychology, (in press)

Marsh, H.W. & Cooper, T.L. Prior subject interest, students' evaluations, and instructional effectiveness. Multivariate Behavioural Research (in press)

Marsh, H.W., Fleiner, H., & Thomas, C.S. Validity and usefulness of student evaluations of instructional quality. Journal of Educational Psychology, 1975, 67, 833-839.

Marsh, H.W., & Overall, J.U. Long-term stability of students' evaluations: A note on Feldman's "Consistency and variability among college students in rating their teachers and courses". Research in Higher Education, 1979, 10, 139-147

Marsh, H.W. & Overall, J.U. Validity of Students' Evaluations of Teaching: A Comparison with Instructor Self Evaluations by Teaching Assistants, Undergraduate Faculty and Graduate Faculty. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April, 1979. (ERIC Document Reproduction Service No. ED177 205) (B)

Marsh, H.W. & Overall, J.U. Validity of students' evaluations of teaching effectiveness: Cognitive and Affective criteria. Journal of Educational Psychology, 1980, 70, 468-475.

Marsh, H.W. & Overall, J.U. Relative influence of course level, course type, and instructor on students' evaluations of instruction. American Educational Research Journal, (in press)

Marsh, H.W., Overall, J.U. & Kesler, S.P. Validity of student evaluations of instructional effectiveness: A comparison of faculty self-evaluations and evaluations by their students. Journal of Educational Psychology, 1979, 71, 149-160 (A)

Marsh, H.W., Overall, J.U. & Kesler, S.P. Class size, students' evaluations, and instructional effectiveness. American Educational Research Journal, 1979, 16, 57-70 (B)

Marsh, H.W., Overall, J.U., & Thomas, C.S. The Relationship Between Students' Evaluation of Instruction and Expected Grade. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, April 1976. (ERIC Document Reproduction Service No. ED 126140).

McKeachie, W.J. Student ratings of faculty: A reprise. Academe, 1979 (October), 384-397.

Overall, J.U. & Marsh, H.W. Midterm feedback from students: Its relationship to instructional improvement and students' cognitive and affective outcomes. Journal of Educational Psychology, 1980, 71 856-865.

Overall, J.U. & Marsh, H.W. Students' evaluations of instruction: A longitudinal study of their stability. Journal of Educational Psychology, 1980, 72, 321-325. (B)

Remmers, H.H. Teaching methods in research on teaching. In Gage (Ed). Handbook on Teaching Chicago: Rand McNally, 1963.

TABLE 1

Factor Analyses of Students' Evaluations of Teaching Effectiveness and the Corresponding Faculty Self Evaluations of Their Own Teaching in All 329 Courses

Evaluation Items (paraphrased)	Factor Pattern Loadings I	II	III	IV	V	VI	VII	VIII	IX
I LEARNING/VALUE									
Course Challenging/Stimulating	42(40)	23(25)	09(-10)	04(04)	00(-03)	15(27)	09(05)	16(23)	29(20)
Learned Something Valuable	53(77)	15(02)	10(-02)	09(04)	01(01)	10(00)	10(04)	17(09)	16(06)
Increased Subject Interest	57(70)	12(05)	08(07)	08(07)	02(-03)	18(08)	03(-04)	19(05)	14(-02)
Learned/Understood Subject Matter	55(52)	12(12)	13(12)	05(03)	03(11)	02(-01)	19(07)	14(-04)	-23(-11)
OVERALL COURSE RATING	36(33)	25(29)	16(09)	12(08)	09(02)	12(16)	13(-08)	14(27)	08(16)
II ENTHUSIASM									
Enthusiastic about Teaching	15(29)	55(42)	16(00)	07(02)	21(15)	10(00)	05(16)	01(09)	05(06)
Dynamic & Energetic	08(03)	70(70)	15(01)	11(06)	08(05)	06(05)	07(16)	01(05)	06(03)
Enhanced Presentations with Humor	10(04)	66(58)	-04(06)	05(01)	13(02)	12(02)	14(07)	02(-18)	-07(-10)
Teaching Style Held Your Interest	09(12)	59(64)	23(20)	16(06)	06(00)	03(14)	10(05)	06(03)	-02(-03)
OVERALL INSTRUCTOR RATING	12(27)	40(54)	23(09)	14(08)	23(02)	11(16)	10(-08)	05(27)	05(16)
III ORGANIZATION									
Instructor Explanations Clear	12(00)	07(24)	55(42)	20(09)	05(04)	10(06)	13(01)	06(23)	-08(-03)
Course Materials Prepared & Clear	06(06)	03(-02)	73(69)	09(01)	10(-02)	09(04)	06(03)	10(03)	01(12)
Objectives Stated & Pursued	19(12)	-05(-08)	49(41)	03(05)	08(05)	14(08)	25(27)	06(05)	06(06)
Lectures Facilitated Note Taking	-03(02)	20(09)	58(53)	-17(07)	-02(05)	14(04)	15(06)	08(01)	-04(-05)
IV GROUP INTERACTION									
Encouraged Class Discussions	04(06)	10(02)	01(03)	84(86)	03(00)	00(00)	06(00)	06(-05)	00(-03)
Students Shared Ideas/Knowledge	02(08)	06(-07)	-04(-01)	85(88)	05(13)	05(01)	08(-02)	08(-10)	-02(01)
Encouraged Questions & Answers	03(-04)	06(09)	14(06)	62(69)	16(-02)	15(03)	07(11)	08(21)	00(01)
Encouraged Expression of Ideas	07(01)	02(06)	01(-11)	73(75)	20(09)	05(07)	09(12)	05(09)	00(-02)
V INDIVIDUAL RAPPORT									
Friendly Towards Students	-04(10)	17(06)	00(-06)	13(12)	68(78)	-01(-05)	13(02)	10(-05)	-07(01)
Welcomed Seeking Help/Advice	04(-10)	05(02)	02(07)	06(00)	85(75)	-04(04)	12(06)	05(20)	03(-04)
Interested in Individual Students	07(10)	11(09)	00(01)	14(07)	69(77)	-01(-09)	14(03)	08(-09)	03(09)
Accessible to Individual Students	02(-13)	-11(-11)	16(09)	09(-02)	62(43)	20(25)	08(13)	00(14)	04(07)
VI BREADTH OF COVERAGE									
Contrasted Implications	-05(02)	12(01)	05(03)	08(01)	-03(01)	72(84)	08(-03)	14(02)	08(-06)
Gave Background of Ideas/Concepts	08(03)	08(10)	16(07)	-03(-02)	02(-02)	71(78)	01(08)	11(-01)	03(03)
Gave Different Points of View	04(-06)	04(09)	11(11)	08(16)	06(01)	72(55)	07(17)	01(-06)	04(08)
Discussed Current Developments	23(29)	08(-04)	-04(-04)	05(12)	09(00)	50(48)	06(05)	16(10)	-01(-02)
VII EXAMINATIONS/GRADING									
Examination Feedback Valuable	-03(01)	08(09)	06(-11)	09(05)	08(12)	-04(03)	72(62)	05(-03)	09(03)
Eval Methods Fair/Appropriate	06(02)	00(-03)	03(14)	07(06)	14(00)	10(17)	69(64)	11(11)	-08(04)
Tested Emphasized Course Content	08(00)	-01(04)	11(21)	01(01)	06(00)	11(-04)	70(58)	07(10)	-02(-03)
VIII ASSIGNMENTS									
Readings/Texts Valuable	-06(09)	-03(-03)	03(07)	-01(-06)	03(01)	07(-07)	01(11)	91(70)	02(04)
Added to Course Understanding	12(01)	-01(-12)	01(04)	09(21)	01(17)	-02(08)	07(05)	81(56)	06(10)
IX WORKLOAD/DIFFICULTY									
Course Difficulty (Easy-Hard)	-06(00)	06(-01)	04(-05)	-04(02)	-01(00)	08(00)	-04(08)	10(04)	85(74)
Course Workload (Light-Heavy)	14(-04)	-09(-01)	03(02)	07(05)	00(04)	06(01)	00(01)	00(-04)	88(86)
Course Pace (Too Slow - Too Fast)	-20(07)	12(00)	04(18)	-12(-09)	06(02)	-03(-07)	04(08)	05(-04)	62(32)
Hours/week Outside of Class	14(00)	07(00)	-11(00)	07(02)	00(02)	-04(03)	03(-08)	05(21)	73(46)

NOTE: Factor loadings in boxes are the loadings for items designed to measure each factor. All loadings are presented without decimal points. Factor analyses of student ratings and instructor self ratings (loadings in parentheses) consisted of a principal-components analysis, Kaiser normalization, and rotation to a direct oblimin criterion. The first nine unrotated factors for the instructor self ratings had eigenvalues of 9.5, 2.9, 2.5, 2.2, 2.0, 1.4, 1.3, 1.1 & 1.0, and accounted for 68% of the variance. For the student ratings the first nine eigenvalues were 19.9, 3.3, 2.3, 1.5, 1.2, 0.9, 0.7, 0.6 & 0.5, and accounted for 98% of the variance. The analyses were performed with the commercially available SPSS routine (See Nie, et. al., 1975).

TABLE 2

Multitrait-Multimethod Matrix: Correlations Between Student and Faculty Self Evaluations in All 329 Courses

INSTRUCTOR SELF EVALUATION FACTORS	INSTRUCTOR SELF-EVALUATION FACTORS								
	LEARN	ENTHU	ORGAN	GROUP	INDIV	BRDTH	EXAMS	ASIGN	WRKLD
LEARNING/VALUE	(83)								
ENTHUSIASM	29	(82)							
ORGANIZATION	12	01	(74)						
GROUP INTERACTION	01	03	-15	(90)					
INDIVIDUAL RAPPORT	-07	-01	07	02	(82)				
BREADTH	13	12	13	11	-01	(84)			
EXAMINATIONS	-01	08	26	09	15	20	(76)		
ASSIGNMENTS	24	-01	17	05	22	09	22	(70)	
WORKLD/DIFFICULTY	03	-01	12	-09	06	-04	09	21	(70)

STUDENT EVALUATION FACTORS	INSTRUCTOR SELF-EVALUATION FACTORS									STUDENT EVALUATION FACTORS								
	LEARN	ENTHU	ORGAN	GROUP	INDIV	BRDTH	EXAMS	ASIGN	WRKLD	LEARN	ENTHU	ORGAN	GROUP	INDIV	BRDTH	EXAMS	ASIGN	WR
LEARNING/VALUE	(46)	10	-01	08	-12	09	-04	08	02	(95)								
ENTHUSIASM	21	(54)	-04	-01	02	-01	-03	-09	-09	45	(96)							
ORGANIZATION	17	13	(30)	-03	04	07	09	00	-05	52	49	(93)						
GROUP INTERACTION	19	05	-20	(52)	00	-02	-14	-04	-08	37	30	21	(98)					
INDIVIDUAL RAPPORT	03	03	-05	13	(28)	-19	-03	-02	00	22	35	33	42	(96)				
BREADTH	26	15	09	00	-14	(42)	00	09	02	49	34	56	17	15	(98)			
EXAMINATIONS	18	09	01	-01	06	-09	(17)	-02	-06	48	42	57	34	50	33	(93)		
ASSIGNMENTS	20	03	02	09	-01	04	-01	(45)	12	52	21	34	30	29	40	42	(92)	
WORKLD/DIFFICULTY	-06	-03	04	00	03	-03	12	22	(69)	06	02	-05	-05	08	18	-02	20	(87

NOTE: Values in the diagonals of the upper left and lower right matrices, the two triangular matrices, are reliability (coefficient alpha) coefficients (See Nie, et. al., 1977). Values in the diagonal of lower left matrix, the square matrix, are convergent validity coefficients that have been corrected for unreliability according to the Spearman Brown equation. The nine uncorrected validity coefficients, starting with Learning would be .41, .48, .25, .46, .25, .37, .13, .36, & .54. All correlation coefficients are presented without decimal points. Correlations greater than .10 are statistically significant.

TABLE 3

Correlations Between 16 Background Variables and 11 Student Evaluation Scores
(N=511 Averages)

STUDENT EVALUATION SCORES

BACKGROUND VARIABLES	Overall Course	Overall Instructor	Learning	Enthusiasm	Organ-ization	Group Interaction	Individual Rapport	Breadth	Exam	Assignments	Workload/ Difficulty
Prior Subject Interest	33	20	44	23	-03	29	09	-03	03	23	12
Workload/Difficulty	23	14	12	06	01	-02	01	15	10	23[a]	xx
Expected Grade	21	20	29	20	01	31	17	-02	18	13	-29
Reason for Taking Course											
% General Interest Only	16	12	15	09	16	07	-02	19[a]	10	18	-13
% Major Elective	16	13	26	06	-03	21	04	18[a]	02	15	-06
% Major Requirement	-15	-12	-18	-07	-08	-04	01	26[a]	-02	-17	17
% General Ed Require	-11	-08	-17	-04	03	-28	-06	03	-09	-06	-12
% Minor/Related Field	07	06	07	03	02	02	04	12	01	07	-01
Class Level											
Course Level	17	14	21	12	-08	29	14	13	04	11	06
% Fresh-Soph in Class	-12	-12	-18	-08	-01	-28	-17	01	-09	-05	-03
% Jr.-Sr. in Class	11	01	07	21	04	05	13	03	10	09	-01
Avg. "Year in School"	11	10	20	11	-06	27	19	-06	03	04	03
Overall GPA (prior)	07	07	10	07	-06	17	14	04	07	13	12
% Division Majors	15	14	15	03	05	29	08	08	13	11	15
Enrollment	-10	-09	-14[a]	01	-03	-32[a]	-18	01	-13	-04	01
Teacher Rank	-02	-08	-10	-12	-10	-14	-05	24[a]	-14	13	11
MULTIPLE R^2 (% Variance Explained)	20%	8.9%	24.7%	5.9%	0%	23.0%	3.5%	11.3%	8.3%	12.3%	19.6%

Note: Correlations are presented without decimal points. Correlations that are underlined indicate background variables which account for at least 5% of the variance in a particular evaluation score. The value of Multiple R squared is based upon the combined effect of the subset of background variables that is most highly correlated with the evaluation score. This was determined with a step-wise multiple regression in which a new background variable was added at each step until no additional variable could increase Multiple R squared by as much as 1%. The Multiple R squared was then corrected for the number of variables in the equation.

[a] These relationships showed substantial non-linearity (i.e., quadratic and/or cubic components add at least 1% to the Variance Explained by the linear relationship and the Total Variance Explained by all components was at least 5%).

USC EVALUATION SERVICES

AS A DESCRIPTION OF THIS COURSE/INSTRUCTOR, THIS STATEMENT IS

(SELECT THE BEST RESPONSE FOR EACH OF THE FOLLOWING STATEMENTS, LEAVING A RESPONSE BLANK ONLY IF IT IS CLEARLY NOT RELEVANT)

		VERY POOR	POOR	FAIR	GOOD	EXCELLENT
1	**LEARNING:** YOU FOUND THE COURSE INTELLECTUALLY CHALLENGING AND STIMULATING	1	2	3	4	5
2	YOU HAVE LEARNED SOMETHING WHICH YOU CONSIDER VALUABLE	1	2	3	4	5
3	YOUR INTEREST IN THE SUBJECT HAS INCREASED AS A CONSEQUENCE OF THIS COURSE	1	2	3	4	5
4	YOU HAVE LEARNED AND UNDERSTOOD THE SUBJECT MATERIALS IN THIS COURSE	1	2	3	4	5
5	**ENTHUSIASM:** INSTRUCTOR WAS ENTHUSIASTIC ABOUT TEACHING THE COURSE	1	2	3	4	5
6	INSTRUCTOR WAS DYNAMIC AND ENERGETIC IN CONDUCTING THE COURSE	1	2	3	4	5
7	INSTRUCTOR ENHANCED PRESENTATIONS WITH THE USE OF HUMOR	1	2	3	4	5
8	INSTRUCTOR'S STYLE OF PRESENTATION HELD YOUR INTEREST DURING CLASS	1	2	3	4	5
9	**ORGANIZATION:** INSTRUCTOR'S EXPLANATIONS WERE CLEAR	1	2	3	4	5
10	COURSE MATERIALS WERE WELL PREPARED AND CAREFULLY EXPLAINED	1	2	3	4	5
11	PROPOSED OBJECTIVES AGREED WITH THOSE ACTUALLY TAUGHT SO YOU KNEW WHERE COURSE WAS GOING	1	2	3	4	5
12	INSTRUCTOR GAVE LECTURES THAT FACILITATED TAKING NOTES	1	2	3	4	5
13	**GROUP INTERACTION:** STUDENTS WERE ENCOURAGED TO PARTICIPATE IN CLASS DISCUSSIONS	1	2	3	4	5
14	STUDENTS WERE INVITED TO SHARE THEIR IDEAS AND KNOWLEDGE	1	2	3	4	5
15	STUDENTS WERE ENCOURAGED TO ASK QUESTIONS & WERE GIVEN MEANINGFUL ANSWERS	1	2	3	4	5
16	STUDENTS WERE ENCOURAGED TO EXPRESS THEIR OWN IDEAS AND/OR QUESTION THE INSTRUCTOR	1	2	3	4	5
17	**INDIVIDUAL RAPPORT:** INSTRUCTOR WAS FRIENDLY TOWARDS INDIVIDUAL STUDENTS	1	2	3	4	5
18	INSTRUCTOR MADE STUDENTS FEEL WELCOME IN SEEKING HELP/ADVICE IN OR OUTSIDE OF CLASS	1	2	3	4	5
19	INSTRUCTOR HAD A GENUINE INTEREST IN INDIVIDUAL STUDENTS	1	2	3	4	5
20	INSTRUCTOR WAS ADEQUATELY ACCESSIBLE TO STUDENTS DURING OFFICE HOURS OR AFTER CLASS	1	2	3	4	5
21	**BREADTH:** INSTRUCTOR CONTRASTED THE IMPLICATIONS OF VARIOUS THEORIES	1	2	3	4	5
22	INSTRUCTOR PRESENTED THE BACKGROUND OR ORIGIN OF IDEAS/CONCEPTS DEVELOPED IN CLASS	1	2	3	4	5
23	INSTRUCTOR PRESENTED POINTS OF VIEW OTHER THAN HIS/HER OWN WHEN APPROPRIATE	1	2	3	4	5
24	INSTRUCTOR ADEQUATELY DISCUSSED CURRENT DEVELOPMENTS IN THE FIELD	1	2	3	4	5
25	**EXAMINATIONS:** FEEDBACK ON EXAMINATIONS/GRADED MATERIALS WAS VALUABLE	1	2	3	4	5
26	METHODS OF EVALUATING STUDENT WORK WERE FAIR AND APPROPRIATE	1	2	3	4	5
27	EXAMINATIONS/GRADED MATERIALS TESTED COURSE CONTENT AS EMPHASIZED BY THE INSTRUCTOR	1	2	3	4	5
28	**ASSIGNMENTS:** REQUIRED READINGS, TEXTS WERE VALUABLE	1	2	3	4	5
29	READINGS, HOMEWORK, ETC. CONTRIBUTED TO APPRECIATION AND UNDERSTANDING OF SUBJECT	1	2	3	4	5
30	**OVERALL:** HOW DOES THIS COURSE COMPARE WITH OTHER COURSES YOU HAVE HAD AT USC?	1	2	3	4	5
31	HOW DOES THIS INSTRUCTOR COMPARE WITH OTHER INSTRUCTORS YOU HAVE HAD AT USC?	1	2	3	4	5

STUDENT AND COURSE CHARACTERISTICS (LEAVE BLANK IF NO RESPONSE APPLIES)

32	COURSE DIFFICULTY: RELATIVE TO OTHER COURSES, WAS (1-VERY EASY ... 3-MEDIUM ... 5-VERY HARD)	1	2	3	4	5
33	COURSE WORKLOAD: RELATIVE TO OTHER COURSES, WAS (1-VERY LIGHT ... 3-MEDIUM ... 5-VERY HEAVY)	1	2	3	4	5
34	COURSE PACE WAS (1-TOO SLOW ... 3-ABOUT RIGHT ... 5-TOO FAST)	1	2	3	4	5
35	HOURS/WEEK REQUIRED OUTSIDE OF CLASS 1: 0 TO 2 ... 2: 2 TO 5 ... 3: 5 TO 7 ... 4: 8 TO 12 ... 5: OVER 12	1	2	3	4	5
36	LEVEL OF INTEREST IN THE SUBJECT PRIOR TO THIS COURSE (1-VERY LOW ... 3-MEDIUM ... 5-VERY HIGH)	1	2	3	4	5
37	OVERALL GPA AT USC? 1: BELOW 2.5 ... 2: 2.5 TO 3.0 ... 3: 3.0 TO 3.4 ... 4: 3.4 TO 3.7 ... 5: ABOVE 3.7 LEAVE BLANK IF NOT YET ESTABLISHED AT USC	1	2	3	4	5
38	EXPECTED GRADE IN THIS COURSE (1=F, 2=D, 3=C, 4=B, 5=A)	F	D	C	B	A
39	REASON FOR TAKING THE COURSE: 1=MAJOR REQUIRE ... 2=MAJOR ELECTIVE ... 3=GENERAL ED REQUIRE ... 4=MINOR RELATED FIELD ... 5=GENERAL INTEREST ONLY ... SELECT THE ONE WHICH IS BEST	1	2	3	4	5
40	YEAR IN SCHOOL: 1: FRESH ... 2: SOPH ... 3: JR ... 4: SR ... 5: GRAD	1	2	3	4	5
41	MAJOR DEPARTMENT: 1=SOC SCI/COMM ... 2=NAT SCI/MATH ... 3=HUMANITIES ... 4=BUSINESS ... 5=EDUCATION ... 6=ENGINEERING ... 7=FINE ARTS ... 8=PUB AFFAIRS ... 9=OTHER ... 10=UNDECLARED/UNDECIDED	1	2	3	4	5
		6	7	8	9	10

SUPPLEMENTAL QUESTIONS (USE RESPONSES BELOW FOR INSTRUCTOR'S QUESTIONS)

42	1	2	3	4	5	47	1	2	3	4	5	52	1	2	3	4	5	57	1	2	3	4	5
43	1	2	3	4	5	48	1	2	3	4	5	53	1	2	3	4	5	58	1	2	3	4	5
44	1	2	3	4	5	49	1	2	3	4	5	54	1	2	3	4	5	59	1	2	3	4	5
45	1	2	3	4	5	50	1	2	3	4	5	55	1	2	3	4	5	60	1	2	3	4	5
46	1	2	3	4	5	51	1	2	3	4	5	56	1	2	3	4	5	61	1	2	3	4	5

INSTRUCTOR'S NAME	DEPARTMENT NAME	COURSE NUMBER

INSTRUCTIONS

This evaluation form is intended to measure your reactions to this instructor and course. Results will be reported to the Department Chairmen to be used as part of the overall evaluation of the instructor. These evaluations will have budgetary and promotional implications so please take it very seriously. When you have finished a designated student will pick up the evaluations and take them to the Department Chairperson. Your responses will remain anonymous and summaries will not be given to the instructor until after the final grades have been assigned.

- ****Put Instructor's Name, Department Name and Course Number at top (i.e., Smith, Psychology, 200)
- ****Use a number 2 pencil, do not use ink, ball point, magic marker, etc.
- ****Blacken only one response for each question and erase any changes completely

OPEN-ENDED COMMENTS

PLEASE INDICATE THE IMPORTANT CHARACTERISTICS OF THIS INSTRUCTOR/COURSE WHICH HAVE BEEN MOST VALUABLE TO YOUR LEARNING EXPERIENCE.

1.

2.

3.

PLEASE INDICATE CHARACTERISTICS OF THIS INSTRUCTOR/COURSE WHICH YOU FEEL ARE MOST IMPORTANT FOR HIM/HER TO WORK ON IMPROVING (PARTICULARLY ASPECTS NOT COVERED BY RATING ITEMS).

1.

2.

3.

PLEASE USE THE ADDITIONAL SPACE TO CLARIFY ANY OF YOUR RESPONSES OR TO MAKE OTHER COMMENTS.

INSTRUCTOR: DOE, JOHN
DEPARTMENT: SAMPLE DEPARTMENT
COURSE: SAMPLE DEPT 999

CLASS SCHEDULE NUMBER: 99999
TERM: SPRING 78

NUMBER OF STUDENTS COMPLETING EVALUATIONS: 43
PERCENTAGE OF ENROLLED STUDENTS COMPLETING EVALUATIONS: 92%

**********THIS IS A SAMPLE OF THE SUMMARY REPORT RETURNED TO INDIVIDUAL FACULTY AND THEIR DEPARTMENT CHAIRPERSONS**********

STUDENT AND COURSE CHARACTERISTICS:
FOR EACH QUESTION THE PERCENTAGE OF STUDENTS MAKING EACH RESPONSE AND THE MEAN AVERAGE RESPONSE (IF APPROPRIATE) IS PRESENTED (THESE STATISTICS ARE BASED UPON THE NUMBER OF STUDENTS ACTUALLY RESPONDING TO THE ITEM). IN ADDITION, THE PERCENTAGE OF STUDENTS WHO COMPLETED THE EVALUATION FORM BUT DID NOT RESPOND TO A PARTICULAR QUESTION IS INDICATED BY THE "NO RESPONSE" PERCENTAGE.

PRIOR INTEREST	PCT
1-VERY LOW	9%
2-	9%
3-MEDIUM	42%
4-	12%
5-VERY HIGH	28%
NO RESPONSE	0%
MEAN: 3.39	

OVERALL G.P.A.	PCT
1-BELOW 2.5	0%
2-2.5 - 3.0	27%
3-3.0 - 3.4	39%
4-3.4 - 3.7	12%
5-ABOVE 3.7	22%
NO RESPONSE	5%
MEAN: 3.28	

EXPECTED GRADE	PCT
0-F	0%
1-D	0%
2-C	14%
3-B	62%
4-A	24%
NO RESPONSE	2%
MEAN: 3.09	

REASON IN CLASS	PCT
1-MAJ REQRD	44%
2-MAJ ELECT	37%
3-GEN ED REQ	2%
4-MIN/RELTD	2%
5-GEN INTRST	14%
NO RESPONSE	0%

YEAR IN SCHOOL	PCT
1-FRESHMAN	9%
2-SOPHOMORE	35%
3-JUNIOR	33%
4-SENIOR	23%
5-GRADUATE	0%
NO RESPONSE	0%
MEAN: 2.69	

MAJOR DEPRTMNT	PCT
1-SOC SCI	72%
2-NAT SCI	2%
3-HUMNTIES	9%
4-BUSINESS	2%
5-EDUCATION	0%
6-ENGINEER	0%
7-PERF ART	0%
8-PUB AFFR	0%
9-OTHER	12%
0-UNDEC	2%

COURSE DIFFICULTY	PCT
1-VERY EASY	2%
2-	2%
3-MEDIUM	40%
4-	40%
5-VERY HARD	16%
NO RESPONSE	0%
MEAN: 3.64	

COURSE WORKLOAD	PCT
1-VERY LIGHT	0%
2-	0%
3-MEDIUM	71%
4-	24%
5-VERY HEAVY	5%
NO RESPONSE	2%
MEAN: 3.32	

COURSE PACE	PCT
1-TOO SLOW	2%
2-	0%
3-RIGHT	71%
4-	19%
5-TOO FAST	7%
NO RESPONSE	2%
MEAN: 3.28	

OUTSIDE HRS/WK	PCT
1-0 TO 2	0%
2-2 TO 5	55%
3-5 TO 7	32%
4-8 TO 12	13%
5-OVER 12	0%
NO RESPONSE	7%
MEAN: 2.56	

SUMMARY EVALUATION SCORES: NINE EVALUATION FACTOR SCORES, TWO OVERALL RATING ITEMS, AND TWO STUDENT/COURSE CHARACTERISTIC ITEMS THE NINE EVALUATION FACTOR SCORES ARE WEIGHTED AVERAGES OF SEPARATE RATING ITEMS AND HAVE A MEAN AVERAGE (ACROSS ALL USC COURSES) OF 50. FOR ALL SCORES, THE STANDARD ERROR (SE) IS A MEASURE OF THE RELIABILITY. IT IS SMALLER (MORE RELIABLE) WHEN LARGER NUMBERS OF STUDENTS ARE RESPONDING AND WHEN THERE IS GREATER AGREEMENT AMONG THE STUDENTS COMPLETING THE EVALUATIONS. DIFFERENCES OF LESS THAN ONE STANDARD ERROR ARE TOO SMALL TO BE RELIABLY INTERPRETED. IN GENERAL, EVALUATIONS BASED UPON LESS THAN 10 STUDENTS' RESPONSES OR LESS THAN 50% OF THE STUDENTS ENROLLED IN THE CLASS SHOULD BE INTERPRETED CAUTIOUSLY. THE PERCENTILE RANKS (WHICH MAY VARY BETWEEN 0 & 100) AND THE CORRESPONDING GRAPHS SHOW HOW YOUR RATINGS COMPARE WITH OTHER COURSES IN YOUR COMPARISON GROUP. HIGHER PERCENTILE RANKS AND MORE STARS INDICATE HIGHER RATINGS. YOUR COMPARISON GROUP IS:
********** UNDERGRADUATE COURSES NOT TAUGHT BY TEACHING ASSISTANTS **********

		MEAN	SE +/-	%TIL RANK	GRAPH OF %TILE RANK 0 1 2 3 4 5 6 7 8 9
	EVALUATION FACTOR SCORES				
LEARNING	VALUABLE LEARNING EXPERIENCE, WAS INTELLECTUALLY STIMULATING/CHALLENGING	58.2	1.745	80	**********
ENTHUSIASM	INSTR DISPLAYED ENTHUSIASM, ENERGY, HUMOR & ABILITY TO HOLD INTEREST	56.9	1.506	74	**********
ORGANIZATION	ORGANIZATION/CLARITY OF EXPLANATIONS, COURSE MATERIALS, OBJECTIVES, LECTURES	61.3	1.726	90	**********
GROUP INTERACT	STUDENTS ENCOURAGED TO DISCUSS, PARTICIPATE, SHARE IDEAS & ASK QUESTIONS	51.3	2.160	53	**********
INDV. RAPPORT	INSTRUCTOR ACCESIBLE, FRIENDLY, AND INTERESTED IN STUDENTS	48.7	2.362	43	**********
BREADTH	PRESENTATION OF BROAD BACKGRD, CONCEPTS & ALTERNATIVE APPROACHES/THEORIES	55.1	2.274	70	**********
EXAMINATIONS	STUDENT PERCEPTIONS OF VALUE & FAIRNESS OF EXAMS/GRADED MATERIALS	50.9	2.188	55	**********
ASSIGNMENTS	VALUE OF ASSIGNMENTS IN ADDING APPRECIATION/UNDERSTANDING TO COURSE	59.9	1.639	88	**********
WORKLOAD/DIFF	RELATIVE COURSE WORKLOAD, DIFFICULTY, PACE, AND OUTSIDE HOURS REQUIRED	51.5	1.515	60	**********
	OVERALL SUMMARY ITEMS				
OVERALL COURSE	HOW DOES THIS COURSE COMPARE WITH OTHERS AT U.S.C.? (QUESTION 30)	4.44	0.102	83	**********
OVERALL INSTR	HOW DOES THIS INSTRUCTOR COMPARE WITH OTHERS AT U.S.C.? (QUESTION 31)	4.61	0.089	84	**********
	IMPORTANT STUDENT/COURSE CHARACTERISTICS				
LEVEL OF INTEREST IN SUBJECT PRIOR TO THIS COURSE (1-VERY LOW ... 5-VERY HIGH) (QUESTION 36)		3.39	0.191	49	**********
EXPECTED GRADE IN THE COURSE (0-F, 1-D, 2-C, 3-B, 4-A) (QUESTION 38)		3.09	0.428	30	*******

PREPARED BY DR. HERBERT W. MARSH, OFFICE OF INSTITUTIONAL STUDIES, USC; L.A., CA 90007 TELEPHONE: (213)741-6503.

INSTRUCTOR: DOE, JOHN
DEPARTMENT: SAMPLE DEPARTMENT
COURSE: SAMPLE DEPT 999

CLASS SCHEDULE NUMBER: 99999
TERM: SPRING 78

NUMBER OF STUDENTS COMPLETING EVALUATIONS: 43
PERCENTAGE OF ENROLLED STUDENTS COMPLETING EVALUATIONS: 92%

EVALUATION ITEMS (SOME QUESTIONS HAVE BEEN ABBREVIATED):

FOR EACH QUESTION, THE PERCENTAGE OF STUDENTS MAKING EACH RESPONSE, THE MEAN AVERAGE RESPONSE, AND THE STANDARD ERROR (SE) OF THE RESPONSES ARE PRESENTED (THESE STATISTICS ARE BASED UPON THE ACTUAL NUMBER OF STUDENTS RESPONDING TO THE QUESTION). IN ADDITION, THE PERCENTAGE OF STUDENTS WHO COMPLETED THE EVALUATION FORM BUT DID NOT RESPOND TO A PARTICULAR QUESTION IS INDICATED IN THE "NO RESP" COLUMN. DIFFERENCES IN MEAN AVERAGES THAT ARE LESS THAN ONE STANDARD ERROR (SEE PAGE ONE FOR A DESCRIPTION) ARE TOO SMALL TO BE RELIABLY INTERPRETED. IN GENERAL, EVALUATIONS BASED UPON LESS THAN 10 STUDENTS' RESPONSES, EVALUATIONS BASED UPON LESS THAN 50% OF THE CLASS, AND EVALUATION ITEMS WHICH WERE FREQUENTLY LEFT BLANK SHOULD BE INTERPRETED CAUTIOUSLY. THE PERCENTILE RANKS (WHICH VARY BETWEEN 0 & 100) AND THE GRAPHS SHOW HOW YOUR EVALUATIONS COMPARE WITH OTHER COURSES IN YOUR COMPARISON GROUP (HIGHER PERCENTILE RANKS AND MORE STARS INDICATE MORE FAVORABLE EVALUATIONS).
YOUR COMPARISON GROUP IS: ********** UNDERGRADUATE COURSES NOT TAUGHT BY TEACHING ASSISTANTS **********

	PERCENTAGE RESPONDING								RANK RELATIVE TO YOUR COMPARISON GROUP (SEE ABOVE)	
	VERY POOR	POOR	MED-IUM	GOOD	VERY GOOD	NO RESP	MEAN	SE +/-	PTIL RANK	GRAPH 0 1 2 3 4 5 6 7 8 9
LEARNING										
1. CRSE WAS INTELLECTUALLY CHALLENGING AND STIMULATING	0	0	5	42	53	0	4.48	0.089	92	**********
2. LEARNED SOMETHING CONSIDERED TO BE VALUABLE	0	0	9	40	51	0	4.41	0.100	88	**********
3. INCREASED INTEREST IN SUBJECT AS CONSEQUENCE OF CRSE	0	2	7	40	51	0	4.39	0.110	82	**********
4. LEARNED AND UNDERSTOOD THE SUBJECT MATERIALS	0	0	9	60	30	0	4.20	0.090	68	**********
ENTHUSIASM										
5. INSTR WAS ENTHUSIASTIC ABOUT TEACHING THE CRSE	0	0	2	26	72	0	4.69	0.077	94	**********
6. INSTR WAS DYNMIC & ENERGETIC IN CONDUCTING CRSE	0	0	7	50	42	2	4.35	0.094	72	**********
7. INSTR ENHANCED PRESENTATION WITH HUMOR	0	5	33	36	26	2	3.82	0.135	42	**********
8. INSTR STYLE OF PRESENTATION HELD INTEREST	0	7	7	53	33	0	4.11	0.125	71	**********
ORGANIZATION										
9. INSTR'S EXPLANATIONS WERE CLEAR	0	0	12	40	49	0	4.36	0.104	80	**********
10. CRSE MATERIALS WERE WELL PREPARED AND EXPLAINED	0	0	7	34	59	5	4.50	0.099	88	**********
11. PROPOSED OBJECTIVES AGREED WITH THOSE ACTUALLY TAUGHT	0	0	5	44	51	5	4.45	0.092	87	**********
12. LECTURES FACILITATED TAKING NOTES	0	2	9	28	60	0	4.46	0.116	90	**********
GROUP INTERACTION										
13. STUDENTS ENCOURAGED TO PARTICIPATE IN CLASS DISCUSSIONS	0	5	19	33	44	0	4.15	0.135	52	**********
14. STUDENTS INVITED TO SHARE IDEAS & KNOWLEDGE	0	2	21	40	37	0	4.11	0.125	46	**********
15. STUDENTS ENCOURAGED TO ASK QUESTIONS AND GIVE ANSWERS	0	0	21	28	51	0	4.29	0.121	59	**********
16. STUDENTS ENCOURAGED TO EXPRESS OWN IDEAS	0	2	21	35	42	0	4.15	0.128	51	**********
INDIVIDUAL RAPPORT										
17. INSTR WAS FRIENDLY TOWARDS INDIVIDUAL STUDENTS	0	5	17	45	32	7	4.04	0.133	26	**********
18. INSTR WELCOMED STUDENTS TO SEEK HELP/ADVICE	0	7	29	29	36	2	3.92	0.149	32	**********
19. INSTR HAD GENUINE INTEREST IN INDIVIDUAL STUDENTS	0	2	25	40	32	7	4.01	0.130	49	**********
20. INSTR WAS ACCESSIBLE DURING OFFICE HRS/AFTER CLASS	0	0	16	50	34	12	4.17	0.111	68	**********
BREADTH										
21. INSTR CONTRASTED IMPLICATIONS OF VARIOUS THEORIES	0	2	19	36	43	2	4.18	0.128	67	**********
22. INSTR PRESENTED BACKGROUND OF IDEAS/CONCEPTS	0	0	9	40	51	0	4.41	0.100	84	**********
23. INSTR PRESENTED POINTS OF VIEW OTHER THAN OWN	0	2	19	45	33	2	4.09	0.121	53	**********
24. INSTR DISCUSSED CURRENT DEVELOPMENTS IN FIELD	0	2	16	35	47	0	4.25	0.125	58	**********
EXAMINATIONS										
25. FEEDBACK ON EXMS/GRADED MATERIALS WAS VALUABLE	2	12	31	40	14	2	3.51	0.148	38	**********
26. METHOD OF EVALUATION WAS FAIR & APPROPRIATE	0	5	21	52	21	2	3.89	0.121	56	**********
27. GRADED MATERIALS TESTED CRSE CONTENT AS EMPHASIZED	0	0	24	40	33	2	4.09	0.116	63	**********
ASSIGNMENTS										
28. REQUIRED READINGS/TEXTS WERE VALUABLE	0	0	7	49	44	0	4.36	0.093	90	**********
29. ASSIGNMENTS CONTRIBUTED TO APPRECIATION/UNDERSTANDING	0	0	10	48	43	2	4.32	0.099	83	**********
OVERALL										
30. HOW DOES THIS COURSE COMPARE WITH OTHERS AT U.S.C.?	0	0	10	36	55	2	4.44	0.102	83	**********
31. HOW DOES THIS INSTRUCTOR COMPARE WITH OTHERS AT U.S.C.?	0	0	5	29	67	2	4.61	0.089	94	**********